HAB (Harmful Algal Bloom) Detection System

COMP47250 Team Software Project Final Report

Daniel Ilyin
School of Computer Science, UCD
Dublin, Ireland
daniel.ilyin@ucdconnect.ie

Sagar Satish Poojary
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Dharmik Arvind Vara
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Karthika Garikapati
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Kruthi Jangam
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Roshan Palem
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Abstract—**Harmful Algal Blooms (HABs) pose serious risks to marine ecosystems, human health, and coastal economies. Yet most traditional detection methods remain slow, expensive, and limited in how much area they can cover. In this project, we developed a smarter, faster alternative: an AI-powered HAB Detection System that uses satellite imagery and deep learning to predict toxic bloom events more effectively.**

At the core of our system is a CNN+LSTM model trained on spatiotemporal data-cubes built from satellite-derived environmental data. The model takes in three key modalities chlorophyll-a concentration, sea surface temperature, and remote sensing reflectance captured across a 10-day window. To make the system accessible to a range of users, we designed a tiered architecture: a lightweight Free Tier using a single modality over 5 days, and more advanced tiers combining richer data for higher accuracy.

Users interact with the system through a simple web interface where they can click on a map to get instant predictions for any location. Our best performing model (Tier 2) achieved 92% accuracy, while the lighter tiers offered reasonable performance with reduced data input and compute needs.

By combining geospatial interactivity, deep learning, and flexible design, this system offers a practical step forward in early HAB detection especially for resource-limited agencies that need timely, reliable insights.

I. INTRODUCTION

Harmful Algal Blooms (HABs) are showing up more frequently and with greater intensity across coastal waters around the world. These toxic blooms, often caused by nutrient runoff and warming ocean temperatures, can do serious damage to marine life, public health, and local economies. In many cases, they've led to large-scale fish deaths, contaminated drinking water, and forced the shutdown of both recreational beaches and fishing operations. Despite the growing threat, most current methods of detecting HABs still rely on manual sampling and lab analysis which are slow, expensive, and not practical for covering large areas in real time.

Our project set out to build a smarter, more scalable approach using satellite data and deep learning. The aim was to design a system that can predict toxic bloom events from space quickly, accurately, and in a way that's easy for people to use. It fits into a larger effort to apply AI to real-world environmental problems, where timely action can help prevent serious harm.

The main challenge we faced was figuring out how to model the way key environmental variables shift over both space and time. To solve that, we used a hybrid model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, trained on spatiotemporal datacubes built from satellite imagery. These datacubes cover 10 days of data across three critical variables: chlorophyll-a levels, sea surface temperature, and remote sensing reflectance.

To make the system flexible and widely usable, we also introduced a tiered prediction structure. Users can choose between a lightweight free version which uses minimal data and more advanced tiers that deliver better accuracy using richer inputs. All of this is wrapped into a simple web interface where users can click on a map and get bloom predictions instantly. The result is a system that combines technical depth with real-world accessibility to stay ahead of harmful algal bloom events.

Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump

II. User Stories and Scenarios

The following scenarios demonstrate how different types of users would interact with our HAB Detection System and shows how our platform fills gaps in current HAB monitoring solutions. Each scenario tries to connect a specific users needs to our technical solution and tiered services, showing quantified improvements over current monitoring approaches.

A. Marine Biologists and Environmental Agencies

1) Current Challenge: Environmental researchers and public health officials require urgent access to comprehensive HAB data for scientific analysis and/or emergency response planning. Traditional monitoring means collecting water samples and waiting 3-5 days for laboratory results which is too slow for effective intervention planning. Additionally coverage is limited to static monitoring stations which are quite sparse leaving most of the water unmonitored.

2) System Interaction: Research institutions and environmental agencies use our Tier 2 services for comprehensive analysis capabilities. Scientists specify coordinates or upload satellite imagery through the dashboard interface, accessing complete 10-day temporal windows across all available modalities (chlorophyll-a, SST, PAR, five RRS bands). The system processes 100km² spatial areas through our CNN-LSTM architecture, providing 94% accurate predictions with sub 30 second response times.

3) Quantified Improvements: This improves HAB monitoring from the traditional days-long workflows into minutes of analysis. Researchers gain access to continuous spatial coverage versus the spare monitoring station coverage from typical monitoring networks.

4) Technical Innovation: This scenario shows our system's ability to convert our spatiotemporal model outputs into fast and accessible, scientifically rigorous data while providing research-grade accuracy standards. The tiered architecture makes sure computational resources scale appropriately as required.

B. Commercial Fishing Operations

1) Current Challenge: Commercial fishermen face huge economic risks from HAB events that can contaminate catches causing costly delays or even damage to equipment. Traditional monitoring provides information far too late to be useful in decision making and doesn't even provide enough spatial and temporal coverage to be useful for plan trips.

2) System Interaction: A commercial fishing operation uses our Tier 1 services for operational planning. Through our dashboard, operators input their fishing coordinates and get 7-day predictions with 80% accuracy. The system processes multi-modal satellite data (chlorophyll-a, SST, PAR) through our CNN-LSTM architecture, providing day-by-day toxicity assessments for their specified fishing areas.

3) Quantified Improvements: This new capability changes their reactive HAB responses into active business preplanning. Operators can schedule fishing activities around predicted safe periods, protecting their equipment investments and catch quality. The extended forecast allows for strategic deployment of vessels and crew, removing any economic losses from unexpected HAB events.

4) Technical Innovation: This use case showcases our ability to deliver operational-grade forecasts with the reliability and temporal resolution required for commercial decision-making, tackling the scalability limitations of traditional monitoring approaches.

C. Recreational Water Users

1) Current Challenge: Members of the public planning recreational activities near water bodies have no easy to use and accessible method to assess HAB risks in advance. Traditional monitoring provides sparse coverage with results available only after laboratory analysis days after sampling.

2) System Interaction: A family planning a weekend lake visit accesses our Free Tier service through the dashboard. Using the interactive map, they select their intended location and visit date. The system processes chlorophyll-a concentration data from recent MODIS imagery, applies our CNN-based classification model and delivers a 5-day toxicity forecast with 70% accuracy within minutes.

3) Quantified Improvements: This provides previously inaccessible satellite data analysis into an intuitive point-and-click user experience. Users receive actionable safety information without requiring any technical expertise, enabling proactive decision-making about their activities. The colour coded risk visualisation (green/yellow/red) allows for easy comprehension of complex environmental data.

4) Technical Innovation: This scenario demonstrates our contribution of providing easy access and understanding to quite complex spatiotemporal modelling. The system converts raw multi-modal satellite data into clear forecasts solving the main challenge of making environmental risk assessment usable to non-technical users.

Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump
Bump

III. State of the Art and Related Work

Harmful Algal Bloom detection has evolved from manual sampling methods to newer machine learning approaches, leading up to HABNet (Hill et al., 2020) [1], which is the current state-of-the-art in HAB predictions. Although its success as a research project HABNet is confined to academic environments due to it's complicated implementation and lack of any user-friendly accessibility.

A. HABNet: Technical Foundation and Critical Limitations

HABNet uses a spatio-temporal deep learning architecture in the form of datacubes, using multiple MODIS modalities (chlorophyll-a, SST, PAR, RRS bands) across 100km×10-day windows [1]. Their hybrid CNN-LSTM system uses NASNet-Mobile for spatial feature extraction followed by LSTM temporal modelling which leads them to achieving a 91% detection accuracy and 86% prediction accuracy up to 8 days ahead.

HABNet outperforms all alternative approaches. Traditional manual sampling achieves $> 95\%$ accuracy but requires multiple days of processing making rapid responses impossible and limiting the usefulness of the information. Remote sensing threshold methods (chlorophyll-a $\geq 20\,\mu g/L$, spectral shape algorithms) provide realtime coverage but have a 30-40% false positive rates. Classic Machine Learning approaches using SVMs achieving 0.56 kappa accuracy [2], Random Forest methods reaching 0.71 kappa accuracy [2], and ensemble approaches achieving 0.86 kappa accuracy [3], but all lack temporal modelling and need manual feature engineering on small datasets.

IV. Implementation

A. System Architecture Overview

V. Evaluation and Results

A. Dataset Description

B. Experimental Setup

C. Detection Performance

D. Prediction Performance

E. System Performance Evaluation

F. User Evaluation

G. Technical Evaluation

VI. Conclusions and Future Work

References

[1] Paul R Hill et al. "HABNet: Machine learning, remote sensing-based detection of harmful algal blooms". In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13 (2020), pp. 3229–3239.

[2] Weilong Song et al. "Learning-Based Algal Bloom Event Recognition for Oceanographic Decision Support System Using Remote Sensing Data". In: Remote Sensing 7.10 (2015), pp. 13564-13585.

[3] B. Gokaraju et al. "Ensemble methodology for improved detection of harmful algal blooms". In: IEEE Geoscience and Remote Sensing Letters 9.5 (2012), pp. 827-831.